Hierarchical Bayesian modeling for combining Dates in archaeological context

Philippe Lanos, Anne Philippe

HIERARCHICAL BAYESIAN MODELING FOR COMBINING DATES IN ARCHAEOLOGICAL CONTEXT

PHILIPPE LANOS[1] AND ANNE PHILIPPE[2]

Abstract

A Bayesian approach is proposed for combining dates from different dating methods used in archeology. This modeling provides an automatic way to penalize outlying data. Examples are provided from different archeological contexts and involving radiocarbon, luminescence and archaeomagnetic results. This new combination procedure is also applied to the wiggle-matching process in dendrochronological dating. Calculations are based on MCMC numerical techniques and can be performed using the cross-platform ChronoModel application which is free, open source software (FOSS).

keywords : Bayesian statistics; ChronoModel software ; individual errors ; MCMC methods ; outlier penalization.

1. Introduction

Many dating methods are available to learn about the age of archeological artifacts. The choice of the method generally depends on the nature of the object and its age (see Aitken (2013) for a detailed review). Radiocarbon, luminescence, and archaeomagnetic dating are probably the most commonly used methods.

Radiocarbon dating (or 14C dating) is used to evaluate the ages of biological specimens (e.g. charcoals, wood artifacts, bone remains, etc.). It is possible to date objects as old as around 50 000 years. Many statistical papers are devoted to this method and software applications are available to analyze and model the radiocarbon data (see Buck et al., 1999; Bronk Ramsey and Lee, 2013; Bronk Ramsey, 2009a).

Luminescence dating (TL/OSL) is based on the radiation absorbed and stored in the crystal lattice of minerals such as quartz. The calculated age is the elapsed time since the last exposure to sunlight or intense heat.

Archaeomagnetic dating (AM) relies on the past variations of the Earth's magnetic field. The ferromagnetic minerals, heated to high temperatures, acquire a remnant magnetization with a direction parallel to, and an intensity proportional to, the local Earth's magnetic field at the time of cooling. The method is implemented for some geographical area where variations of the Earth's magnetic field are known. It allows the last heating of baked clays to be dated.

One important question in archeology is the estimation of the date of an archeological event that is associated with a set of contemporaneous artifacts. This problem comes up frequently when constructing chronologies in archeology. Each artifact of the event is dated, and so we collect a

Date: December 10, 2015

[1] CNRS IRAMAT-CRPAA, Université Bordeaux-Montaigne and Géosciences-Rennes, Université Rennes 1.

[2] Laboratoire de mathématiques Jean Leray, Unversité Nantes.

sample of observations provided by one or several dating methods. This leads to the issue of combining measurements in order to define the event date.

The simplest statistical model is defined as n independent measurements $M_1, \ldots, M_n$ assumed to have the same unknown mean μ. The parameter μ is in turn related to calendar time via so-called calibration curve. We denote as s_i^2 the experimental variance on each measurement M_i, that is supposed known and evaluated by the laboratory during the measurement process. To summarize, we write

$$M_i = \mu + s_i \epsilon_i, \qquad \forall\, i = 1, ..., n \tag{1}$$

where $\epsilon_1, ..., \epsilon_n$ are n independent identically distributed Gaussian random variables with zero mean and variance 1. The maximum likelihood estimate of μ (Ward and Wilson, 1978) is given by

$$\hat{\mu}_n = \frac{\displaystyle\sum_{i=1}^{n} \frac{M_i}{s_i^2}}{\displaystyle\sum_{i=1}^{n} \frac{1}{s_i^2}}. \tag{2}$$

The measurements M_i with high variance s_i^2 are penalized, and thus they contribute less to the estimation of the mean μ. In Galbraith et al. (1999), this model is named the "common age model" and is applied for combining paleodoses in the luminescence dating method. This method is also implemented for combining 14C ages (see Bronk Ramsey, 2009a).

To take into account individual effects, we can add a random effect on the parameter μ as follows:

$$\begin{aligned} M_i &= \mu_i + s_i \epsilon_i \\ \mu_i &= \mu + \sigma \lambda_i, \end{aligned} \tag{3}$$

where $\lambda_1, ..., \lambda_n, \epsilon_1, ..., \epsilon_n$ are $2n$ independent and identically distributed Gaussian random variables with zero mean and variance 1. Error $\sigma\lambda_i$ represents the uncertainty between the measurements and the event μ due to sampling (representativeness) problems of unknown origin which are not related to the measurement process in the laboratory. This model is named the "central Age model" in Galbraith et al. (1999). An explicit form of the likelihood estimate is not available. Alternatively, a Bayesian approach can be adopted to estimate the parameters μ and σ^2 of this hierarchical model. It is thus necessary to choose a prior distribution on (μ, σ^2). This choice is discussed for instance in Congdon (2010) or Spiegelhalter et al. (2004) in the particular case of meta-analysis.

In archeology, dating laboratories might provide measurements for

- a 14C age,
- a paleodose measurement (TL/OSL),
- an inclination, a declination or an intensity of the geomagnetic field (AM).

The laboratory measurements are then converted into calendar dates using a calibration curve (see Section 2). This step must be added to the combination model. We thus propose to extend the model with the random effect defined in (3) by adding a calibration step, and including individual effects on the variance by replacing σ by σ_i.

The inclusion of individual effects is motivated by the fact that each measurement can be affected by irreducible errors (Christen, 1994) which can come from different sources such as:

(1) The way of ensuring that the samples studied can realistically provide results for the events that we wish to characterize (measurement or date)
(2) The care taken in sampling in the field,
(3) The care taken in sample handling and preparation in the laboratory,
(4) Other non-controllable random factors that can appear during the process.

The rest of the paper is organized in three sections. Section 2 is devoted to the Bayesian calibration of measurements and the combination of dates. In Section 3, we describe our hierarchical Bayesian model for combining measurements in the context of dating problems and define an "event model" to estimate a date for contemporary artifacts. Section 4 provides an application to wiggle-match dating of tree-ring sequences.

Remark 1. *We define each Bayesian model using a directed acyclic graph (DAG). This graph describes the dependencies in the joint distribution of the probabilistic model. Each random variable of the model (that is an observation or a parameter) appears as a node in the graph. Any node is conditionally independent of its non-descendents given its parents. Nodes represented by circles correspond to the random variables of the model. With the color of the circles, we distinguish between observations (red) and parameters (blue). The green squares indicate exogenous variables.*

2. Calibration and linked models

In this section we describe two standard Bayesian models:

- the individual calibration of a measurement provided by a dating laboratory,
- the combination of measurements performed on the same artifact using the same dating method.

2.1. Calibration.

The simplest problem is the calibration of an individual measurement M in calendar time t. The measure M has a Gaussian distribution with mean μ and known variance s^2. Following a hierarchical model, the variable μ has a Gaussian distribution with mean $g(t)$ and variance $\sigma_g^2(t)$, where g is a function called "calibration curve" linking the measurement to calendar time t. Every dating method has its own calibration curve. For instance, the curve IntCal13 (Reimer et al., 2013) converts a radiocarbon Age in calendar date for the Northern Hemisphere atmospheric samples. The calibration of 14C ages is detailed by Bronk Ramsey (2009a); Buck et al. (1996); Litton and Buck (1995). In archaeomagnetism the secular variation curves of the Earth's magnetic field are used to convert an inclination, declination or intensity measurement to a calendar date (see Lanos, 2004). In the case of an age M provided by luminescence (TL or OSL), the calibration function is merely a linear function $g(t) = (t_0 - t)$ where t_0 is the determination year of the age measurement M in the laboratory.

The construction of the calibration curve depends on the dating method. In the context of OSL dating the calibration curve comes from a physical equation, and so σ_g is assumed to be null. In radiocarbon or archaeomagnetic dating methods, the curve comes from an estimation step (e.g. a non linear regression on a set of dated reference measurements). In this case σ_g corresponds to the errors due to the estimation method. Note that these curves are regularly updated according to the improvement of the reference data.

The Bayesian model for 'calibration' or conversion of a measurement M to calendar date t is the following. The model on the measurement M is given by

$$M = \mu + s\epsilon,$$

where μ is the true value of the measurement and s^2 is the known variance. The distribution of ϵ is the standard Gaussian distribution.

The calibration step converts μ to calendar date t, using the relation

$$\mu = g(t) + \sigma_g(t)\rho,$$

where both functions g and σ_g are assumed known, and where ρ is a standard Gaussian random variable.

Let T be the time range of interest. This interval is fixed from prior information on the historical period to which the event belongs.

We choose the uniform distribution on T as prior distribution on the parameter t,

$$p(t) \propto 1_T(t). \tag{4}$$

If we integrate the posterior distribution of (t, μ) with respect to latent variable μ, we get the posterior distribution of t (up to a multiplicative constant):

$$p(t|M) \propto \frac{1}{S(t)} \exp\left(\frac{-1}{2S^2(t)}(M - g(t))^2\right) 1_T(t), \tag{5}$$

where

$$S(t)^2 = s^2 + \sigma_g^2(t). \tag{6}$$

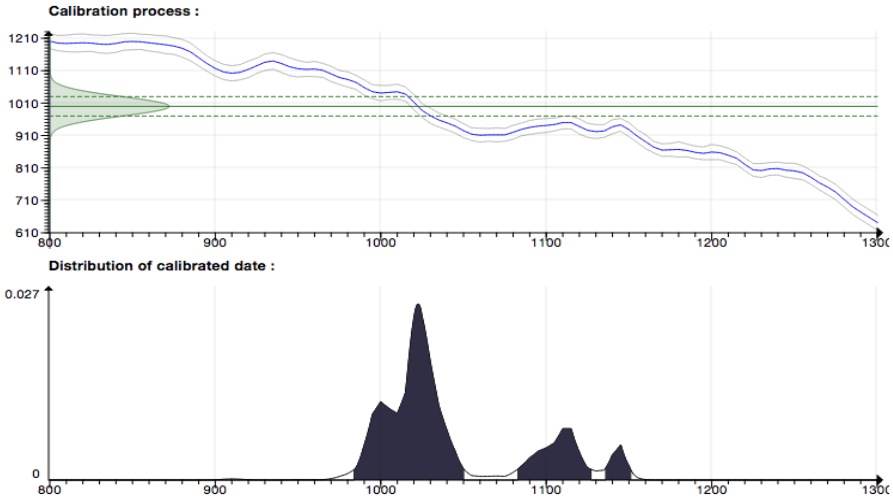

FIGURE 1. Conversion of a 14C age ($A = 1000 \pm 30$ BP) to a calendar date via the calibration curve IntCal13.

Figures 1, 2 and 3 illustrate the calibration process for different calibration curves. In many cases, the individual calibration provides an estimation of the date with a high uncertainty. The posterior distribution is often multi-modal or unimodal with large variance. This motivates the archeologist to repeat measurements on different contemporaneous artifacts using one or several dating methods.

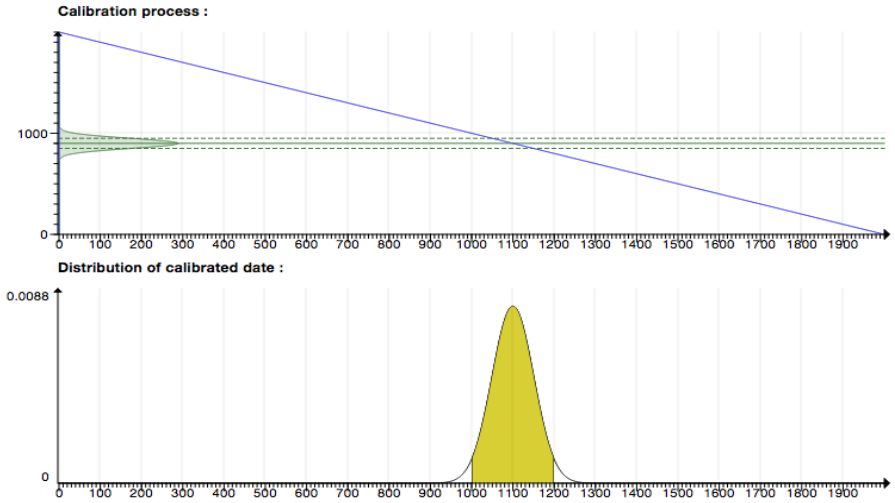

FIGURE 2. Conversion of a TL age ($A = 900 \pm 50$) to a calendar date through the linear transformation $g(t) = (2000 - t)$.

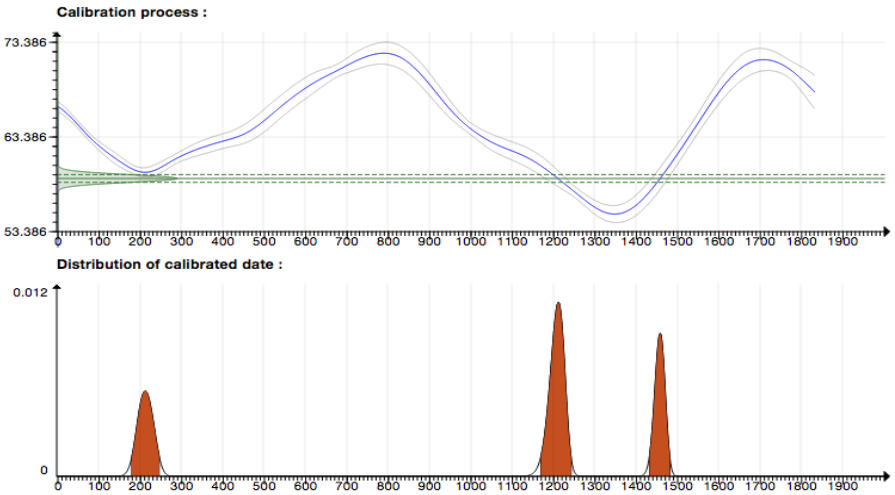

FIGURE 3. Conversion of an inclination measurement ($I = 59 \pm 1$) to a calendar date via the calibration curve of archaeomagnetic field in France (Paris) over the last two millennia.

2.2. **Calibration from multiple measurements.** We observe m independent measurements M_k performed on the same object. We assume that all the measurements can be calibrated with the same curve (g, σ_g^2). This appears, for example, in the case of radiocarbon dating when the same object is analyzed by different laboratories. In this case, the measurements have to be combined before calibration (see Ward and Wilson (1978) and Bronk Ramsey (2009a)).

The Bayesian model (Fig. 4) is defined by the following distributions:

$$\begin{aligned} M_i|\mu &\sim \mathcal{N}(\mu\ ,\ s_i^2) \qquad \forall\ i = 1, ..., m \\ \mu|t &\sim \mathcal{N}(g(t)\ ,\ \sigma_g^2(t)) \\ t &\sim \text{Unif}(T), \end{aligned}$$

where T is the time range of interest. This model is called R-Combine in Bronk Ramsey (2009a). One can easily get the marginal posterior density of t, which is given by

$$p(t|M_1,...,M_m) \propto \frac{1}{\tilde{S}_m(t)} \exp\left(\frac{-1}{2\tilde{S}_m(t)^2}(\overline{M}_m - g(t))^2\right) 1_T(t),$$

where

$$\tilde{S}_m(t)^2 = \bar{s}_m^2 + \sigma_g^2(t) \qquad \text{with} \qquad \frac{1}{\bar{s}_m^2} = \sum_{k=1}^{m} \frac{1}{s_k^2},$$

and where

$$\overline{M}_m = \sum_{k=1}^{m} \frac{M_k}{s_k^2} \Big/ \sum_{k=1}^{m} \frac{1}{s_k^2}.$$

The R-combine model is equivalent to the individual calibration of the observation $(\overline{M}_m, \bar{s}_m^2)$ using the process described in Section 2.1.

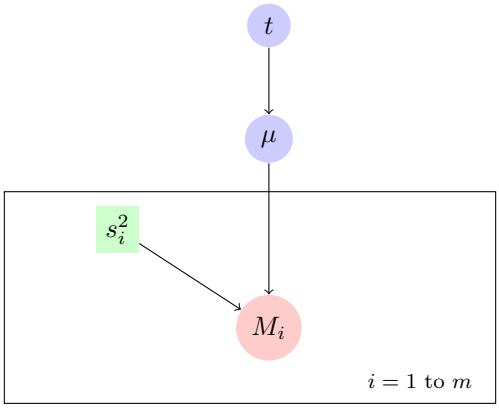

FIGURE 4. DAG of the R-Combine model.

3. EVENT MODEL

In the general case we observe n measurements M_i such that each measurement provides a dating through a calibration step defined by a calibration function g_i and its error σ_{g_i}. The R-combine model is no longer valid in this case because it requires a common calibration curve (i.e. $g_i = g$ for all $i = 1,...,n$). Our idea is to adapt the Bayesian combination of measurements to estimate the date of an archeological event from the individual dating of contemporaneous artifacts.

3.1. The model.

We describe the so-called Event model to estimate a date θ from n measurements M_i provided by different dating methods. We assume that each measurement M_i is related to an individual date t_i through a calibration curve g_i. Here this curve is assumed known with some known uncertainty. The main assumption in our Event model is the contemporaneity of the dates t_i, $i = 1...,n$ with the event date θ. We assume that $\theta \in T$ where T is the bounded interval corresponding to the range of study.

In this context the model with random effect given by (3) can be rewritten as follows

$$\begin{aligned} M_i &= \mu_i + s_i\epsilon_i, \\ \mu_i &= g_i(t_i) + \sigma_{g_i}(t_i)\rho_i, \\ t_i &= \theta + \sigma_i\lambda_i, \end{aligned} \tag{7}$$

where $(\epsilon_1, ...\epsilon_n, \rho_1, ..., \rho_n, \lambda_1, ..., \lambda_n)$ are independent and identically distributed Gaussian random variables with zero mean and variance 1.
The random variables $(\lambda_i)_i$ and $(\epsilon_i)_i$ are interpreted as follows :

- $\sigma_i\lambda_i$ represents the irreducible error between t_i and θ due to sampling problems external to the laboratory (Section 1),
- $s_i\epsilon_i + \sigma_{g_i}(t_i)\rho_i$ represents the experimental error provided by the laboratory and the calibration step.

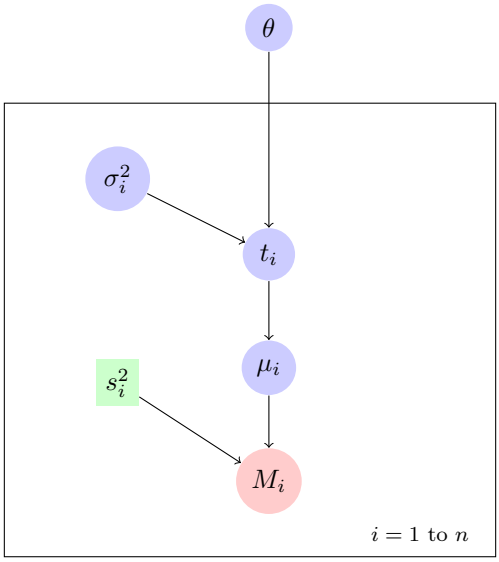

FIGURE 5. DAG for the hierarchical Event model applied to dating combination with calibration.

According to the DAG defining the event model (Fig. 5), the joint distribution of the probabilistic model can be written in the form

$$p(M_1, ..., M_n, \mu_1, ..., \mu_n, t_1, ..., t_n, \sigma_1^2, ..., \sigma_n^2, \theta) = p(\theta) \prod_{i=1}^{n} p(M_i|\mu_i)p(\mu_i|t_i)p(t_i|\sigma_i^2, \theta)p(\sigma_i^2), \tag{8}$$

where the conditional distributions that appear in the decomposition are given by

$$M_i|\mu_i \sim \mathcal{N}(\mu_i\ ,\ s_i^2)$$
$$\mu_i|t_i \sim \mathcal{N}(g_i(t_i)\ ,\ \sigma_{g_i}^2(t_i))$$
$$t_i|\sigma_i^2, \theta \sim \mathcal{N}(\theta\ ,\ \sigma_i^2) \tag{9}$$
$$\sigma_i^2 \sim \text{Shrink}(s_0^2) \tag{10}$$
$$\theta \sim \text{Unif}(T). \tag{11}$$

The density of the uniform shrinkage distribution with parameter s_0^2 (denoted Shrink(s_0^2)), that appears in (10), is given by

$$p(\sigma_i^2) = \frac{s_0^2}{(s_0^2 + \sigma_i^2)^2}. \tag{12}$$

Note that this form of density implies that the random variable $s_0^2/(\sigma^2+s_0^2)$ is uniformly distributed on $[0, 1]$. This distribution is rather diffuse in the sense that the variance and mean are infinite. So it can be considered weakly informative. The properties of the uniform shrinkage prior and

the choice of s_0^2 are discussed in (Daniels, 1999). In the particular case of the central age model defined in (3), Spiegelhalter et al. (2004) suggests the following choice of parameter:

$$\frac{1}{s_0^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{s_i^2}.$$

Parameter s_0^2 quantifies the magnitude of error on the measurements. As s_0^2 is the median of the uniform shrinkage prior, this choice ensures that the parameter σ_i^2 has the same prior probability to be smaller or larger than s_0^2. Therefore we do not favor the measurement errors with respect to the error between t_i and θ due to sampling problems.

With calibrated measurements, the parameter s_0^2 cannot be calculated directly from the variances s_i^2. Indeed the measurement errors s_i are not necessarily homogeneous units (see for instance the case of archaeomagnetism combined with 14C). Moreover they do not contain the information on the variance of the combined variables. To adapt this strategy we have to estimate the variance of the dates t_i. We propose to estimate these variances after the step of individual calibration. We proceed in the following way, for each $i = 1, ..., n$:

(1) An individual calibration step is performed for each measurement M_i
   (a) compute the posterior distribution of t_i (using (5))
   (b) approximate the posterior variance $w_i^2 = \mathrm{var}(t_i|M_i)$
(2) Take as shrinkage parameter s_0:

$$\frac{1}{s_0^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_i^2}.$$

Remark 2. *For the hyperparameters σ_i^2 in the third stage of the hierarchical model, the choice of a diffuse noninformative prior may be problematic as improper priors may induce improper posteriors. This problem appears in our model, for instance by taking $p(\sigma_i) \propto 1/\sigma_i^2$, that is the classical non informative prior (called the Jeffreys prior) for Gaussian observations. The choice of diffuse priors in such models raises particular issues, discussed for instance in Congdon (2010). It is always possible to choose a proper probability distribution which approximates the diffuse prior, for instance*

- *a density proportional to the Jeffreys prior truncated on an closed interval $[a, b] \subset]0, \infty[$,*
- *an inverse Gamma distribution with small scale and shape parameters.*

However, such priors may cause identifiability problems as the posteriors are close to being empirically improper. Moreover the inference of the parameter of interest is sensitive to the choice of parameters in the proper prior distribution. This is a major drawback for small samples.

3.2. **MCMC algorithm.** The posterior distribution of the parameter of interest θ can not be obtained explicitly. It is necessary to implement a computational method to approximate the posterior distribution, its quantiles, the Bayes estimates and the highest posterior density (HPD) regions. We adopt a MCMC (Markov Chain Monte Carlo) algorithm known as the Metropolis-within-Gibbs strategy because the full conditionals cannot be simulated by standard random generators. For each parameter, the full conditional distribution is proportional to (8).

1. For the parameter of interest θ we identify a truncated Gaussian distribution on the period of interest T. Such a distribution can be simulated using rejection sampling. Different choices of proposal distribution are possible, such as the Gaussian distribution or the truncated Laplace (double exponential) distribution on T. It is also possible to use an adaptive random walk Metropolis-Hastings (MH) with a Gaussian proposal distribution.

2. The density of the full conditional distribution of σ_i^2 is explicitly computable up to an unknown multiplicative constant, but it is not a standard distribution. Therefore it is simulated using an adaptive random walk MH with a Gaussian proposal distribution on the variable $\log(\sigma_i^2)$.
3. The full conditional distribution of t_i, $i = 1, ..., n$, is proportional to

$$\frac{1}{S_i(t_i)} \exp\left\{\frac{-1}{2S_i^2(t_i)}(M_i - g_i(t_i)^2\right\} \exp\left\{\frac{-1}{2\sigma_i^2}(t_i - \theta)^2)\right\},$$

where $S_i(t_i)$ is defined in (6). We can choose an adaptive random walk MH with a Gaussian proposal. However, the random walk solution is not necessarily the most efficient choice because the target distribution can be multimodal. We are frequently confronted with this problem as shown by, for instance, the results yielded by archaeomagnetic dating in Example 2. In this context, an alternative is to choose an independent Hasting-Metropolis algorithm with a proposal distribution that mimics the individual calibration density defined in (5). This ensures that all the possible values of t_i can be visited when the calibrated date distribution is multimodal.

These algorithms are implemented in the cross-platform ChronoModel application (Lanos et al., 2015; Vibet et al., 2015), which is free and open source software. In the examples discussed below, the graphics summarizing the numerical results are performed using ChronoModel software. We represent the marginal densities of

- the parameter of interest θ (on gray background)
- the individual measurements t_i (on white background)
- the individual standard deviations σ_i

For each density, the bar above the density represents the shorter 95% posterior probability interval. The vertical lines, delimiting the colored area under the density curve, indicate the endpoints of the 95% HPD region. For a unimodal posterior density, the 95% posterior probability regions coincide.

Different graphical tools are implemented to assess the convergence of the MCMC in ChronoModel software (e.g. the autocorrelation functions, the acceptance rate of Metropolis-Hasting algorithms). The user can adjust the length of burn-in, the maximum number of iterations for adaptation and for acquisition, and the thinning rate. Moreover multiple chains can be simulated and exported for further analysis using, for instance, the R package `coda`.

In term of computational performance, among the examples given in this paper, the longest computing time is achieved for Ex. 2, Indeed 10^6 iterations are required to well estimate the multimodal posterior distribution of the Event. The simulation is performed on a 2 GHz Intel Core i7 by using a single core. The computational time is less than 1 minute including the post-processing i.e. estimation of HPD regions, credibility regions, posterior densities, etc from simulated Markov chain, and displaying results.

3.3. **Examples.**

Example 1. Mont-Saint-Michel (Normandy, France)

An archeology building study has been conducted at the Mont-Saint-Michel site in Normandy, France (Blain et al., 2007; Sapin et al., 2008). The aim of this study was to estimate the dates of different building states of the Carolingian church Notre-Dame-sous-Terre (NDST) located inside the Mont-Saint-Michel abbey. Therefore it is reasonable to fix the time range T equal to [0 , 2000], which widely includes the Carolingian period.

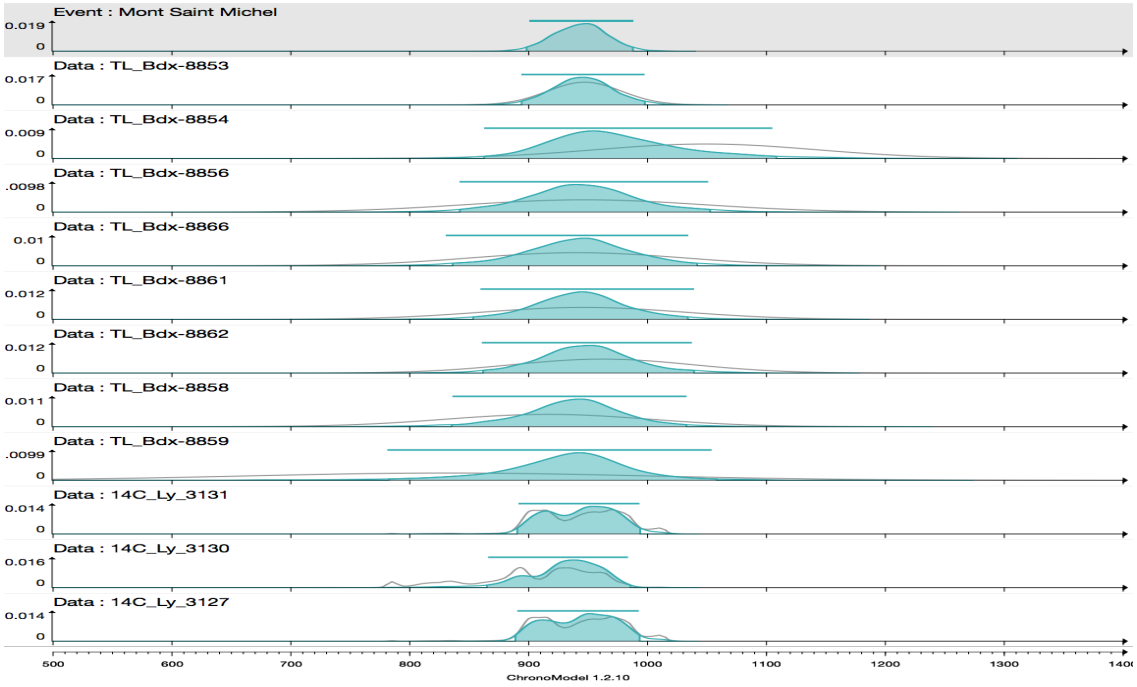

FIGURE 6. Mont-Saint-Michel, Ex. 1. Posterior density of the event date θ (gray background). Posterior densities of t_i (blue line & white background) and individual posterior calibrated densities (gray line & white background) obtained for TL dates and for ^{14}C dates.

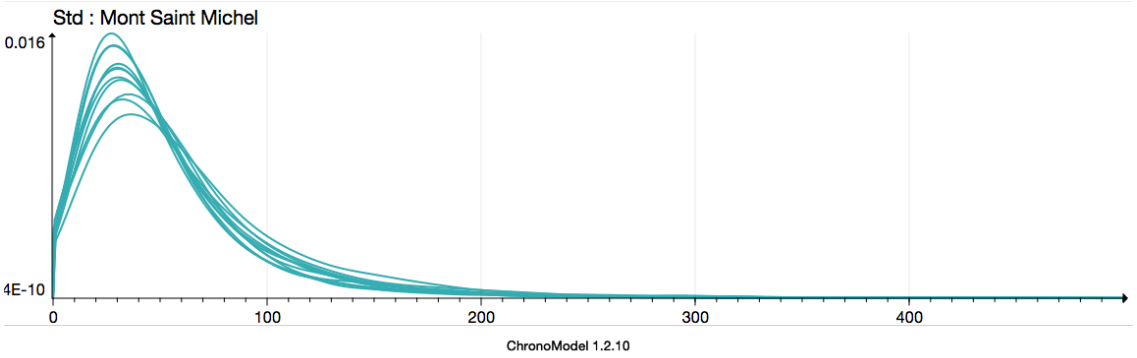

FIGURE 7. Mont-Saint-Michel, Ex. 1 (Cont.). Posterior densities obtained for standard deviations σ_i.

In what follows, we model building state number 1 which is associated with 14C analysis on 3 charcoals, and luminescence analyses of 8 bricks.

Figures 6 and 7 summarize the estimation results obtained by the event model. The 95% HPD interval for the date of building state 1 is [892 ; 993]. This result is more precise than the dating obtained in Blain et al. (2007), which is based solely on luminescence analysis. This result confirms the value of combining different dating methods.

Figure 6 provides a comparison between the individual calibrations of each measurement and the posterior densities of the dates t_i. Clearly the event model improves the precision of the dating

of each artifact included in the event. The graphical comparison of the posterior densities of t_i confirms our assumption that the bricks and charcoals are contemporary.

Posterior densities for standard deviations σ_i fall within the small range $[0, 200]$ at 95% (Fig. 7). These results confirm that the dates t_i are contemporary

Example 2. Cuers (Provence, France), medieval or modern lime kiln

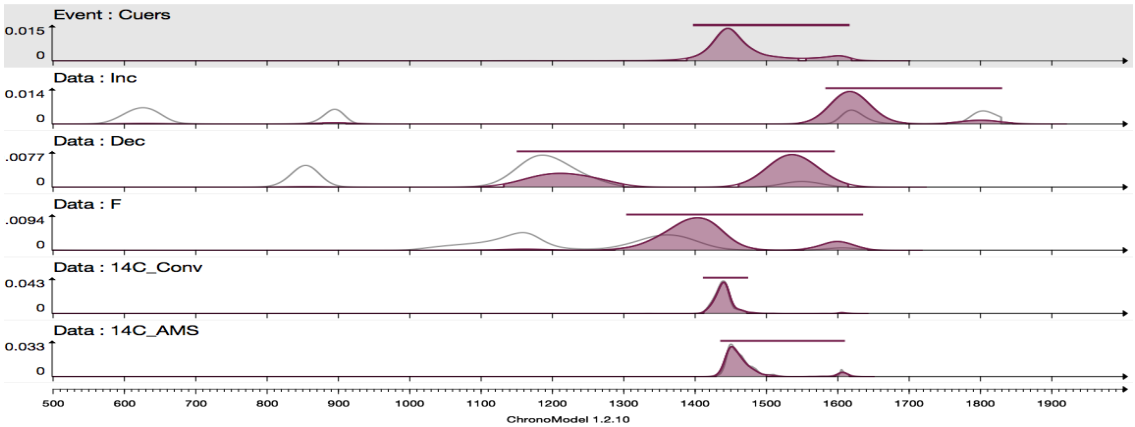

FIGURE 8. Cuers, Ex. 2. Posterior density of the event date θ (gray background). Posterior densities of t_i (red line & white background) and individual posterior calibrated densities (gray line & white background) obtained for archaeomagnetic and ^{14}C dates.

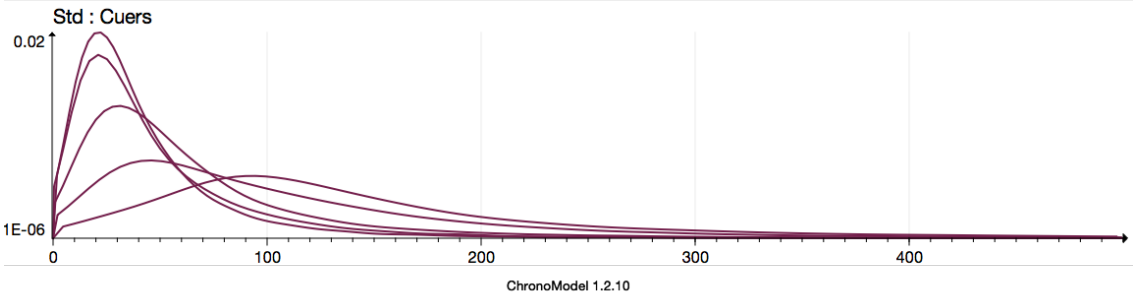

FIGURE 9. Cuers, Ex 2 (Cont.) Posterior densities obtained for standard deviations σ_i.

Archeological excavation was carried out to estimate the last firing date of a lime kiln in Cuers (Provence, France) (Vaschalde et al., 2014). We choose the time range T equal to [0 , 2000]. Indeed no historical or archeological information is available about this site between the Roman period and the 20th century.

The studied artifacts are pieces of baked clay from the kiln wall submitted for AM analysis and two charcoals submitted for 14C analysis. The obtained measurements are three archaeomagnetic parameters (inclination, declination and intensity) and two radiocarbon ages. Figure 8 gives the results obtained by the event model.

As in the previous example, posterior densities t_i (in red) are much more precise than the individual calibrated densities, in the sense that the number of local modes is reduced. The Event

model gives a 95% HPD region for the last firing date which is the union of two disjoint intervals $[1388, 1548]$ and $[1556, 1616]$. In this case the dating is indeterminate in the sense that two periods remain possible for the last firing. Consequently, additional measurements or prior information are required to decide between these two solutions.

Figure 9 gives the posterior densities of standard deviations σ_i. It clearly shows that this parameter takes higher values in the case of archaeomagnetic dating. This indicates a lack of consistency between the dates within the event. The archaeomagnetic dates are penalized in the estimation of the event. This is explained by the fact that radiocarbon dating provides more accurate dates after the calibration step.

Example 3. Tell Qasile, context X (Israel)

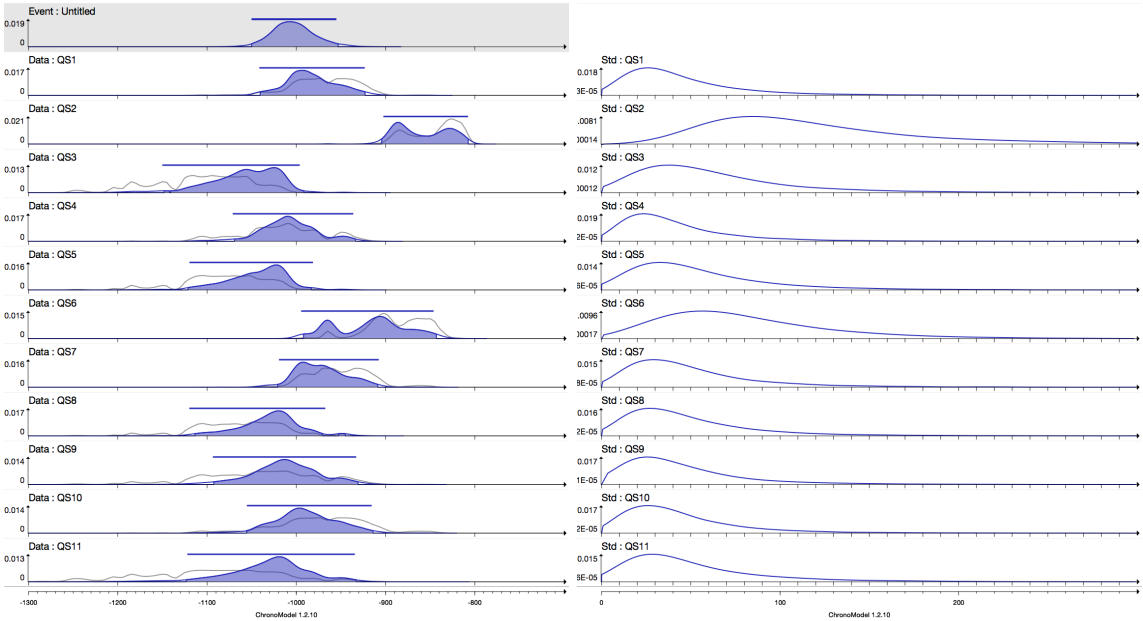

FIGURE 10. Tell Qasile, Ex. 3. [left] Posterior density for Event date θ (gray background). Posterior densities t_i for ^{14}C dates (blue line and white background) The individual posterior calibrated densities are superimposed in gray. [right] Posterior densities obtained for standard deviations σ_i.

We consider the example studied in Bronk Ramsey (2009b), in order to compare the robustness to outliers of the event model with the R-combine model with the outlier implemented in OxCal software.

From eleven radiocarbon dates, the aim is to determine the date of the context X in Tell Qasile, an archeological site in Tel Aviv. Since the site belongs to the protohistoric period, we fix the time range T equal to $[-2000\ ,\ 0]$.

In OxCal, it is necessary to fix the prior probability that each measurement is an outlier. The recommended value is 5% when no pertinent information is available. The 95% HPD region for the date of the context X yielded by OxCal is $[-1054; -970] \cup [-962; -934]$.

The event model gives a uni-modal posterior density and 95% HPD interval equal to $[-1050; -951]$ (Fig. 10). This is very similar to the two merged OxCal intervals.

Posterior densities for standard deviations σ_i are very similar (Fig. 10) except for samples QS2 and QS6 which show higher posterior values, and thus appear to be outliers. The same two outliers are also detected by OxCal.

In conclusion, the event model is robust to outliers and has the advantage that it does not require prior information about the outliers. The ability of the individual variance σ_i^2 to take large values automatically penalizes an outlier.

4. Wiggle-matching

The "wiggle-matching" model combines radiocarbon dating and dendrochronology. Radiocarbon dating is carried out on tree-ring samples separated by a known number of tree-rings. This gives prior information on the calibrated dates t_i. The number of years between these dates is then known, (Manning et al., 2010; Galimberti and Ramsey, 2004; Christen and Litton, 1995).

In this case, the calibrated dates t_i should be shifted according to their relationship to the event date θ. So we adapt our event model to wiggle-matching as follows.

We consider that the event date θ corresponds to the date of a chosen reference tree-ring (for instance the oldest tree-ring, see Example 4).

We denote by δ_i the number of years between t_i and θ, which is fixed by counting the number of rings between them and assuming that one ring is equal to one year. By convention, δ_i is positive (resp. negative) when t_i is older (resp. younger) than θ.

Equation (7) can be rewritten as follows

$$t_i = \theta - \delta_i + \sigma_i \lambda_i. \tag{13}$$

Remark 3. *It is also possible to assume δ_i unknown in order to model a counting error in the number of tree-rings. This extension of the model is implemented in chronomodel software.*

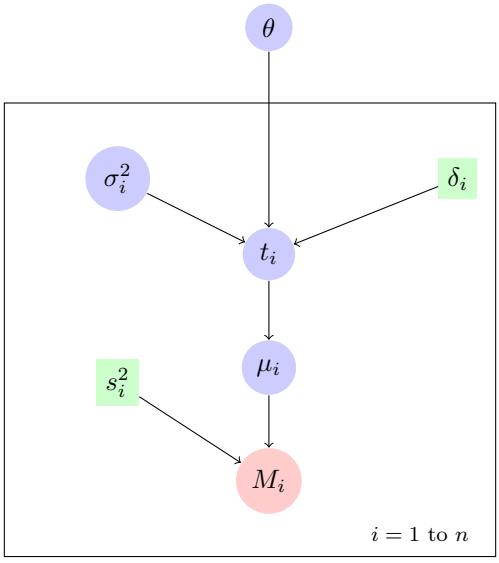

Figure 11. DAG for the wiggle matching model.

Example 4. Gordion Juniper dendrochronology (Central Anatolia)

An application of the Event wiggle-matching model is presented here in the case of the Gordion Juniper dendrochronology (Manning and Kromer, 2011, table 1). Thirty-five samples of this sequence, from relative tree-rings centered at 776.5-1025.5 have been dated by the Heidelberg radiocarbon laboratory. The time range T is fixed equal to $[-5000\ ,\ 0]$, the lack of prior information leads to the choice of a very large interval.

The samples are calibrated with IntCal04.14c curve (Reimer et al., 2004) and are separated with known gaps taking values between 1 and 19 years over a range of 249 years. δ_i is the gap between the dated tree-ring and reference. Following Manning and Kromer (2011), we fix as reference, the oldest tree-ring i.e. $\delta_1 = 0$.

The model considered in Manning and Kromer (2011) is implemented in OxCal, by using the D-Sequence and Gap functions with the outlier model. It gives a 95% HPD interval equal to [-1734, -1724].

The Event model gives a 95% HPD interval for θ equal to [-1744, -1719], which is less precise than the estimate yielded by OxCal. The difference in the 95% HPD region is due to the event structure, which increases the variance of the posterior distribution. However our approach brings robustness. As an illustration, we artificially modify the dataset by adding outliers with the same experimental variances on the 14C ages (see Table 1). Figure 12 gives the estimation obtained by the event model for:

- the original Gordion dataset analyzed in Manning and Kromer (2011),
- the same dataset contaminated with 11 outliers.

TABLE 1. Gordion Juniper dendrochronology (Central Anatolia). Contaminated samples in Gordion dataset

sample number with Manning and Kromer (2011) notations	modified 14C age (BP)
20144, 27605	3100
20137, 25792, 20155	3600
20157, 20141, 20147, 25793, 20153, 27612	3700

Figure 12 shows that the posterior distribution of the event date is very weakly sensitive to the presence of these outliers. Indeed, on the contaminated sample, the event model gives an 95% HPD interval for θ equal to [-1749 ; -1717]. The two results are quite similar; the event date is not affected by the presence of the outliers. The event model considered in Manning and Kromer (2011) cannot be estimated using OxCal software because of poor agreement indices. This alert message indicates in particular that the convergence of the MCMC is not guaranteed.

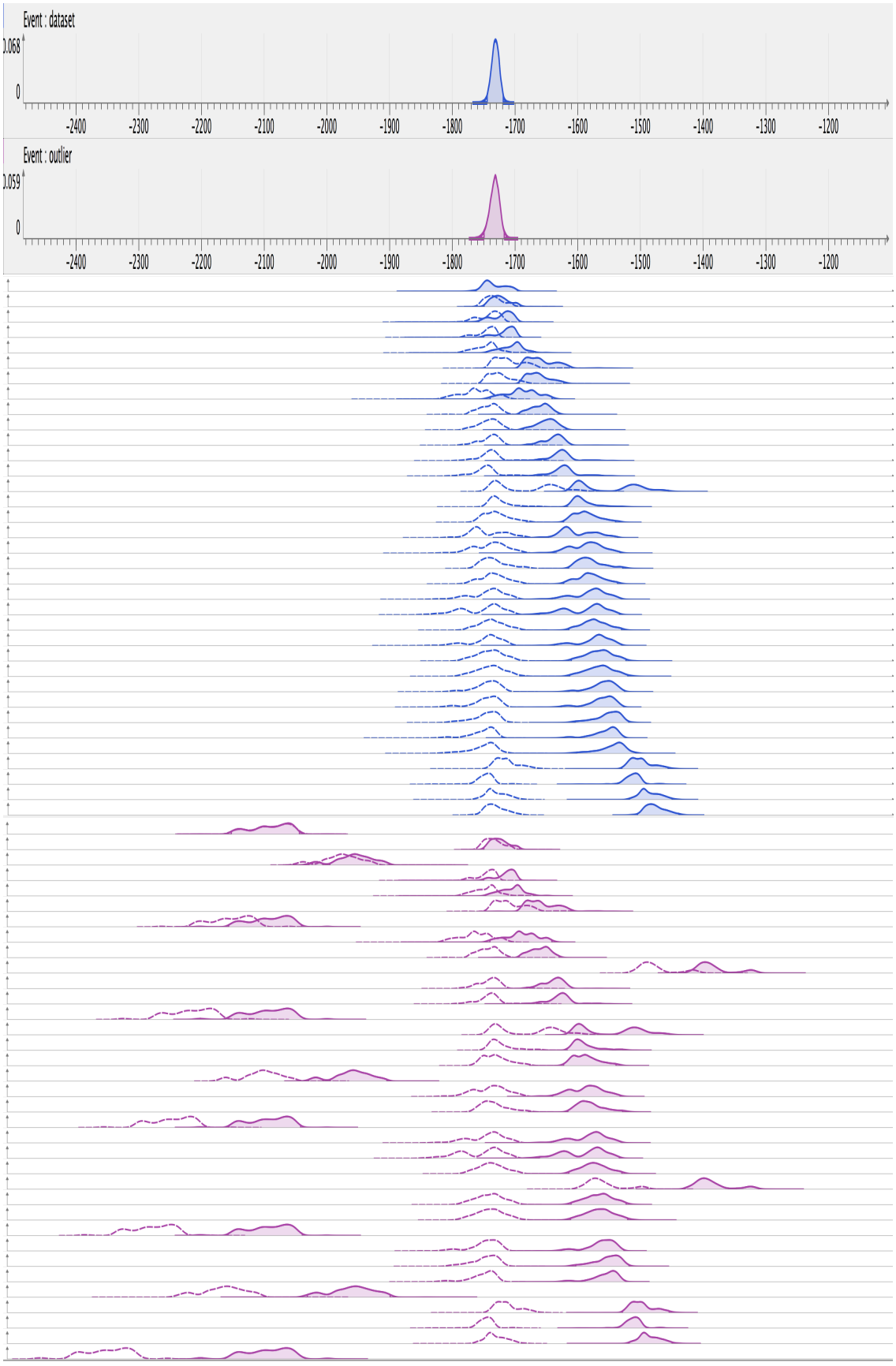

FIGURE 12. Gordion juniper dendrochronology (Central Anatolia). Posterior densities of the dates t_i (solid line and white background), the shifted dates $t_i + \delta_i$ (dotted line and white background) and the Event date θ (gray background) obtained on original Gordion dataset (blue) and on a contaminated version with 11 outliers defined in Table 1 (in red).

Perspective

A combination procedure to estimate the date of an archeological event from a set of contemporaneous artifacts using a hierarchical Bayesian approach is proposed. The event model defines a fundamental chronological entity that can be related to other entities in various ways to produce complex chronological models. We can then construct chronologies of events which take into account new archeological information such as ordering constraints based on stratigraphy, phasing with succession or with a hiatus between phases, or with a duration constraint on the phase, etc. These features are implemented in the ChronoModel software.

Acknowledgement

The authors are extremely grateful to Thomas S. Dye for his careful reading and his insightful comments that helped to improve the exposition in the manuscript.
This project is supported by the grant ANR-11-MONU-007 ChronoModel.

References

Aitken, M. J. (2013). *Science Based Dating in Archaeology*. Routleedge.

Blain, S., Guibert, P., Bouvier, A., Vieillevigne, E., Bechtel, F., Sapin, C., and Baylé, M. (2007). TL-dating applied to building archaeology: The case of the medieval church Notre-Dame-Sous-Terre (Mont-Saint-Michel, France). *Radiation Measurements*, 42(9):1483–1491.

Bronk Ramsey, C. (2009a). Bayesian analysis of radiocarbon dates. *Radiocarbon*, 51(1):337–360.

Bronk Ramsey, C. (2009b). Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon*, 51(3):1023–1045.

Bronk Ramsey, C. and Lee, S. (2013). Recent and planned developments of the program OxCal. *Radiocarbon*, 55:720–730.

Buck, C., Christen, J., and James, G. (1999). BCal : an on-line Bayesian radiocarbon calibration tool. *Internet Archaeology*, 7.

Buck, C. E., Litton, C. D., and G., C. W. (1996). *The BAyesian Approach to Interpreting Archaeological Data*. Chichester, J.Wiley and Son, England.

Christen, J. and Litton, C. (1995). A Bayesian approach to wiggle-matching. *Journal of Archaeological Science*, 22(6):719–725.

Christen, J. A. (1994). Summarizing a set of radiocarbon determinations: a robust approach. *Applied Statistics*, 43(3):489–503.

Congdon, P. D. (2010). *Applied Bayesian hierarchical methods*. CRC Press, Boca Raton, FL.

Daniels, M. J. (1999). A prior for the variance in hierarchical models. *Canad. J. Statist.*, 27(3):567–578.

Galbraith, R. F., Roberts, R. G., Laslett, G. M., Yoshida, H., and Olley, J. M. (1999). Optical dating of single and multiple grains of quartz from Jinmium rock shelter, Northern Australia: Part i, experimental design and statistical models. *Archaeometry*, 41(2):339–364.

Galimberti, M. and Ramsey, C. B. (2004). Wiggle-match dating of tree-ring sequences. *Radiocarbon*, 46(2):917–924.

Lanos, P. (2004). Bayesian inference of calibration curves: Application to archaeomagnetism. In Buck, C. and Millard, A., editors, *Tools for Constructing Chronologies*, volume 177, pages 43–82. Lecture Notes in Statistics, Springer-Verlag.

Lanos, P., Philippe, A., Lanos, H., and Dufresne, P. (2015). *Chronomodel : Chronological Modelling of Archaeological Data using BAyesian Statistics. (Version 1.1) http://www.chronomodel.fr.*

Litton, C. D. and Buck, C. E. (1995). The Bayesian approach to the interpretation of archaeological data. *Archaeometry*, 37(1).

Manning, S., Kromer, B., Ramsey, C. B., Pearson, C., Talamo, S., Trano, N., and Watkins, J. (2010). 14C record and wiggle-match placement for the Anatolian (Gordion Area) Juniper tree-ring chronology 1729 to 751 cal BC, and typical Aegean/Anatolian (growing season related) regional 14C offset assessment. *Radiocarbon*, 52(4).

Manning, S. W. and Kromer, B. (2011). Radiocarbon dating archaeological samples in the eastern mediterranean, 1730 to 1480 bc: further exploring the atmospheric radiocarbon calibration record and the archaeological implications. *Archaeometry*, 53(2):413–439.

Reimer, P., Bard, E., Bayliss, A., Beck, J., Blackwell, P., Ramsey, C. B., Buck, C., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P., Guilderson, T., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T., Hoffmann, D., Hogg, A., Hughen, K., Kaiser, K., Kromer, B., Manning, S., Niu, M., Reimer, R., Richards, D., Scott, E., Southon, J., Staff, R., Turney, C., and van der Plicht, J. (2013). Intcal13 and marine13 radiocarbon age calibration curves 0-50,000 years cal bp. *Radiocarbon*, 55(4).

Reimer, P. J., Baillie, M. G. L., Bard, E., Bayliss, A., Beck, J. W., Bertrand, C. J. H., Blackwell, P. G., Buck, C. E., Burr, G. S., Cutler, K. B., Damon, P. E., Edwards, R. L., Fairbanks, R. G., Friedrich, M., Guilderson, T. P., Hogg, A. G., Hughen, K. A., Kromer, B., McCormac, G., Manning, S., Bronk Ramsey, C., Reimer, R. W., Remmele, S., Southon, J. R., Stuiver, M., Talamo, S., Taylor, F. W., van der Plicht, J., and Weyhenmeyer, C. E. (2004). Intcal04 terrestrial radiocarbon age calibration, 0-26 cal kyr bp. *Radiocarbon*, 46(3):1029–1058.

Sapin, C., Bayle, M., Buttner, S., Guibert, P., Blain, S., Lanos, P., Chauvin, A., Dufresne, P., and Oberlin, C. (2008). Archeologie du bati et archeometrie au Mont-Saint-Michel, nouvelles approches de Notre-Dame-Sous-Terre. *Archeologie Medievale*, 38:71–122.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *BAyesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, Chichester.

Vaschalde, C., Hervé, G., Lanos, P., and Thiriot, J. (2014). La datation des structures de cuisson: intégration de l'archéomagnétisme et du radiocarbone, apports de l'anthracologie. *Archéologie Médiévale*, 44:1–16.

Vibet, M.-A., Philippe, A., Lanos, P., and Dufresne, P. (2015). Chronomodel user's manual. www.chronomodel.fr.

Ward, G. K. and Wilson, S. R. (1978). Procedures for comparing and combining radiocarbon age determinations: a critique. *Archaeometry*, 20(1):19–31.